Report on Potential Scientific Impact and Design Considerations and Objectives for the Proposed Facility for the Production of Proteins and Molecular Tags (GTL-1)

Prepared by a Subcommittee of the DOE Biological and Environmental Research Advisory Committee (BERAC)

December 2004

A Biological and Environmental Research Advisory Committee (BERAC) Subcommittee met on October 11, 2004, to consider the questions in regard to the BER GTL Program and facility-1 (GTL-1) as requested by DOE Office of Science Director, Dr. Raymond Orbach in his letter dated September 23, 2004, to Dr. Keith Hodgson, Chairman of BERAC. The members of the Subcommittee are listed at the end of the report and a copy of the charge letter is also included.

EXECUTIVE SUMMARY

Unequivocally, the GTL-1 facility is essential to meet the scientific goals of the GTL program, which are to empower and transform the world of microbial biology and to enable biological research to further specific missions of the DOE Office of Science in areas of energy, environmental remediation and global climate change. It is only through the use of high-throughput, economical approaches that we can capitalize on the dramatically growing volume of new genome information, enabling and accelerating the essential step of studying and understanding microbial function on the cellular to molecular level. Only then will the potential of utilizing microbial biology as a key strategy in addressing DOE missions be realized. The success of such an endeavor will be far reaching to our Nation's economy and quality of life for decades to come.

The GTL-1 protein production facility is the crucial first step in the process designed to attain the goals of the GTL program. By providing high-throughput production capabilities to make proteins and related affinity reagents and tags, this GTL-1 facility will enable the scientific community to work at the cutting edge of investigating and understanding protein function. The Subcommittee met to reply to a number of specific questions posed with regard to the structure, function and site choice for the facility.

Summary conclusions:

- The GTL-1 facility should initially target expression, production and characterization of the proteins from the genes of several well-studied model microorganisms, such as *E. coli*, *Bacillus subtilis,* and *Saccharomyces cerevisiae* (yeast) that have broad user communities and extensive experimental and data foundations. An early focus of the GTL-1 facility towards this effort will have a transforming effect on gaining a system-level understanding of these organisms. Such comprehensive investigations of model organisms will dramatically aid research on establishing a solid foundation for undertaking the GTL-driven production goals.

- The effort should shift as expediently as possible to the primary goal of studying the DOE mission-related microorganisms and processes, building on experience gained with the model systems.

- The GTL-1 facility is appropriate in scope and goals as developed and described in the most recent documentation that was made available to the Subcommittee for review.

- The development and production components of GTL-1 are most effectively located in a central facility. However, the research and development components that will generate new technologies and solutions to problems need to be distributed to take advantage of all possible avenues of innovation, broadly engaging the academic, national laboratory and industrial research communities. The R&D component of the facility will be critical to its success and long term viability.

- The central facility needs to be housed in a specially designed building that will have highly flexible space that will accommodate evolving high-throughput GTL science 10-20 years from now.

- The facility needs certain large, specialized instrumentation, including mass spectrometers, fermentation facilities, extensive robotics, clean rooms and large computing capacity. However, location next to a synchrotron light source is not considered to be necessary.

- Understanding protein function is of great interest and importance to biological and biomedical research. The GTL-1 protein production facility will serve a large user community and complement efforts supported by NIH and NSF. Technology developments will be shared to the benefit of all.

SUBCOMMITTEE PROCESS AND DETAILED DISCUSSION

The Subcommittee received a package of information on the GTL program and proposed facilities to review in advance of its meeting on October 11, 2004. The meeting began with overview talks on the GTL program and GTL-1 facility presented by Dr. Ari Patrinos and Dr. Jim Fredrickson. The Subcommittee then went into closed session and discussed the scope and need for the GTL-1 facility and how it would serve the goals of GTL. The questions posed in the charge to the Subcommittee were explicitly considered and discussed. Consensus was reached on all of the points. The specific questions posed in the letter from Dr. Orbach were:

1. Why is this facility essential to meeting the scientific goals of the program?
2. Is the facility described the right one in terms of output/capabilities/dimensions to meet the goals?
3. Does it have to be a central facility or can it be distributed?
4. As described, does it require a specially designed building or can the needs be met with leased space, equipment, etc.?
5. What scientific probes, such as synchrotron light sources or mass-spectrometers, are required to make the facility successful?
6. Does it have to be next to a light source or not?

GOALS OF THE GTL PROGRAM AND RELATIONSHIP TO GTL-1

Key overarching goals of the GTL program (http://doegenomestolife.org/) are:

Goal 1: Identify and characterize the molecular machines of life — the multiprotein complexes that execute cellular functions and govern cell form

Goal 2: Characterize gene regulatory networks

Goal 3: Characterize the functional repertoire of complex microbial communities in their natural environments at the molecular level

Goal 4: Develop the computational methods and capabilities to advance understanding of complex biological systems and predict their behavior

The GTL-1 protein production facility is the crucial first step for attainment of the GTL goals, as it is required to provide large numbers of well-characterized proteins and specialized reagents for testing interactions, for building metabolic networks, and for developing affinity and micro array methods to examine the complexity of protein expression in diverse microbial communities and in diverse environments. The capabilities of GTL-1 will enable scientists to focus on challenging problems rather than performing repetitive production activities. The GTL program will need to work with the GTL research teams to prepare them for science that will make optimum use of the output of this facility and drive its initial targets.

By providing high-throughput, economical, global protein production capabilities, this GTL-1 facility will enable the scientific community to work at the cutting edge of investigating and understanding protein function in the framework of the GTL program goals. The availability of well characterized proteins and other reagents important to their study (*e.g.*, affinity reagents, proteins modified with tags) produced by the GTL-1 facility will make a powerfully enabling and transforming contribution to the GTL science program. It will also offer significant advances to the larger scientific community in ways that are complementary to the protein production goals of other agencies like NIH.

The initial production focus for this facility should be on a select portfolio of related microbial organisms and protein families chosen to have optimal scientific impact on the DOE BER GTL science program and its goals. In the first five years of operation, the primary activity should be to develop, test and fully implement the pipeline for protein expression, production and characterization for selected model microbes. These microbes represent classes of organisms that will contribute to the science and technology missions of the facility in important and specific ways:

1. Organisms that are sufficiently well characterized and have large, established user communities such that they are effective models for other microbial biological systems and are well suited to evaluating and optimizing GTL-1 facility capabilities. These include *E. coli, B. subtilis and S. cerevisiae.*

2. Organisms that are already being worked on that are of direct interest to the DOE mission such as *Shewanella*, *Geobacter,* and *Desulfovibrio.*

3. Organisms that should be worked on because they are representative of important large classes of microorganisms such as *Pyrococcus,* a well-studied Archaea for which robust genetic systems are available, and photosynthetic organisms such as *Prochloroccus, Rhodobacter, Synechocystis* and *Rhodopseudomonas,* which are important to DOE's energy and environmental missions.

The strategy for the GTL-1 facility should also feature selection of specific families of important proteins that derive from many organisms across broad evolutionary or environmental regimes. For example, great benefits will result from producing and characterizing many examples of cytochromes, stress-related proteins and rhodopsins. This approach will not only give the maximum information about these particular protein families but will also allow scientists to take advantage of "experiments" of evolution to find the best tools for particular applications. As knowledge of these organisms and protein families progresses, additional new organisms and protein families can be added.

The challenge of providing complete coverage of the proteome of a microorganism is currently beyond the state of the art technology. Some genes have evolved to produce only very small amounts of protein products. Most proteins are idiosyncratic with respect to conditions for production and purification. Some proteins are not readily soluble. Some are relatively unstable and require discovery of special conditions for their care. Others will only function in a properly reconstituted assembly and may need to be produced in such a form. This activity will be an especially crucial precursor to achieving the goals of the future GTL-2 facility for Characterization and Imaging of Molecular Machines. We cannot even be sure that we can identify all the DNA sequences in a genome that should encode proteins. Yet this Subcommittee fully embraces and endorses comprehensiveness as the ultimate goal for the GTL-1 facility in order to push technology development and to advance far beyond the proteins that are most easily expressed. This situation is analogous to the evolution of the Human Genome Project, in which early catalogs of cDNA sequences and then partial draft genome sequences made very valuable contributions to the advance of science long before a full reference genome sequence was achieved. These early products were essential technical steps and precursors on the road to a final, high quality reference genome sequence. In the case of proteomics as will be explored by the GTL science program and enabled by the proteins produced by GTL-1, we know that the chemical diversity of the proteome, the difficulty of knowing all the expressible open reading frames, and the wealth and breadth of fundamental biological questions that will benefit from this program mean that the asymptotic approach to completeness will be longer and more difficult than it was for DNA sequence. Nevertheless, a successful systematic, high-throughput centralized effort will have tremendous impact and it will drive us toward completeness.

FACILITY CAPABILITIES

State of the art technologies, though far from meeting all the challenges of this facility's mission, are ready to be integrated into a robust, centralized high-throughput facility with the ability to:

1. Produce expressible clones for the maximum possible fraction of the proteins encoded by the microbial genomes of choice. These clones should be made available to the user community.

2. Screen conditions for expression and purification of proteins using high-throughput technologies. Characterize all purified proteins, also using high-throughput technologies, and provide them to users. Develop or adopt new technologies to improve expression and purification that can be applied across the proteome. This will include, for example, the need to co-express some proteins and to deal with membrane protein expression, proper folding, co-factor incorporation, etc. Such proteome-wide technologies do not currently exist.

3. Make all protocols and data for individual proteins readily available to the user community.

The management of the facility must have as a constant goal the timely integration of new technology into the facility in order to maximize the effectiveness of production. Clearly, development of technologies within the facility should be encouraged to keep the group at the cutting edge, but it is equally important to support distributed technology development to take advantage of all possible avenues of innovation including academic, national laboratory and industrial communities.

The facility must be very responsive both to needs of the users and to new, emerging technologies. The management must have a style that focuses on deliverables and goals and that is mindful of assuring that new technologies are robust and ready for production.

While the principal goal will be on microbes of specific interest to DOE missions, there should be close cooperation with the biomedical research community to take advantage of communal research and development opportunities – talks with other agencies are already underway.

The ability of the facility to evolve is crucial. Technologies and approaches that would be used today are not the ones that will be best when the facility opens.

The facility will need the following capabilities:

Production line: current technology and design considerations

The process starts with the DNA sequence (available from an annotated genome sequence database) coding for a particular protein. A computer algorithm selects the best set of methods/protocols for producing the specific protein, and then automatically executes pre-experimental steps such as designing and ordering primer oligonucleotides and scheduling pre-production screening. DNA is prepared using an automated oligonucleotide synthesizer, various ligation steps, and purification procedures; DNA clones are either prepared *de novo* or selected from the cryogenic archive; or the DNA sequence information is decoded to provide the amino acid sequence for use in the chemical synthesis process.

Synthesized DNA

The sequence is purified of contaminating chemicals; amplified using a PCR machine and the sequence is verified by sequencing in an automated DNA sequencer. The DNA preparation can then be used directly in the *in vitro* coupled transcription and translation system to produce protein or it can be cloned into an appropriate vector for use in an *in vivo* expression system. The samples are then ready for production screening.

Expression

The clones are transfected into one or more expression systems and appropriate clones are selected by an automated colony picker and the sequence is verified by sequencing in an automated DNA sequencer, and the cultures are then ready to be grown for preproduction screening.

Preproduction Screen

High-throughput trials will test different conditions to maximize production of proteins and their purification in a native form. Protein production, purity, and solubility are determined by a combination of light scattering, mass spectrometry, and UV absorbance. Data are analyzed automatically by a computer program, which also selects the best conditions/methods for a production run.

It is expected that the capability to produce large numbers of clones and express and purify a large number of proteins in very high-throughput, cost effective ways will greatly enable production of some of the most difficult proteins to produce because a large number of different clones and expression and purification conditions can be explored.

High-Throughput Protein Production

Using the selected conditions, expressing cells will be grown in larger amounts (prokaryote or eukaryote) in a multi-sample fermentor. Proteins will be extracted and purified using liquid chromatography, capillary electrophoresis, or affinity columns. The primary focus will be on producing modest amounts of each protein: 1-10mg. However, for some proteins that might be needed in larger quantities, the facility should also have the capability to produce hundreds of milligrams.

In addition to the established approaches, two alternative methods should be considered for producing proteins of choice:

Protein production using cell-free systems. DNA can also be used as a template in cell-free extracts that transcribe RNA from the DNA and then translate the RNA into protein. This technology, though not yet highly developed for economical and effective high-throughput, may evolve into a system of choice.

Chemical synthesis of proteins. Solid phase synthesis of peptides can be carried out using well established systems but currently there are length limitations that need substantial improvement if this is to become routine.

Characterization - Biophysical Properties of Proteins

An array of complementary biophysical and biochemical techniques needs to be used to assess the quality of protein preparations and to provide initial insight into their physical properties beyond simple amino acid sequence (*i.e.*, beyond their primary structure). Also critical is an analysis of the degree of heterogeneity of the protein sample due to post translational and other modifications.

These techniques address five main areas:

1. Activity when available, to ensure the protein produced is biologically useful and meaningful.
2. Protein secondary structure and native state of the protein (Fourier Transform Infrared Spectroscopy (FT-IR), Ultraviolet Circular Dichroism Spectroscopy (UV-CD).
3. Tertiary structure (Fluorescence Emission/Lifetime (FIE/L) and FT-IR).
4. Quaternary structure and particle shape (size exclusion chromatography coupled with laser light scattering (SEC-LLS), small angle x-ray or neutron scattering (SAXS/SANS), Electron Microscopy (EM).
5. Protein stability as measured by changes in the fluorescence or UV-CD properties of the protein.

Each of these analytical techniques provides important information concerning the structure/stability or state of a protein but each also has limitations. By combining these different techniques, it should be possible to gain basic information that will be very helpful to the users. Thus, an integrated suite of protein characterization tools is necessary.

These measurements must be capable of high-throughput and be cost effective. Much of the needed instrumentation is laboratory based (*i.e.,* can be located within the GTL-1 facility). There are some measurements that could benefit from a facility like a high brightness synchrotron or neutron source. For example, at such a possibly remote synchrotron facility, high-throughput systems (flow or robotic-enabled) could be developed and evaluated as a means to provide a cost effective platform for making certain types of valuable measurements on protein samples (*e.g.,* small angle x-ray scattering (SAXS) or extended range circular dichroism (or UV-CD). Results of such developments could be evaluated for their utility in the context of GTL-1 production goals. To take advantage of such an approach, methods would need to be developed for transporting and automated sample handling, data logging/processing and comparison of the results obtained by these methods. The results would need to be integrated with other laboratory-based measurements.

In addition to physical/structural analyses, it is envisioned that high-throughput assays for specific biochemical functions and sensitivities pertinent to DOE applications will be critical. For example, proteins thought to catalyze a subset of enzymatic processes of interest to DOE missions (*e.g.*, hydrogen production, metal reduction, *etc.*) will be assayed for enzymatic activity or metabolite binding properties.

Each of these steps and processes will have to deal with very large numbers of biological samples that need to be tracked appropriately through the automated

systems. There will be a great need for sophisticated bioinformatics analysis at all steps. The processes will generate vast amounts of valuable data on the clones, the proteins and their characterization that need to be properly captured and disseminated to the scientific user community. Implementation of appropriate LIMS systems and data mining capabilities will be absolutely crucial to achieving high-throughput cost-effective clone and protein production as well as to enabling the use of these materials to contribute to the goals of GTL and the Department of Energy. This will require large computing resources and the development of the best scientific tools to properly mine the invaluable data being produced.

Affinity Reagent Production Line

Tagged protein reagents will need to be prepared through the high-throughput protein production line using appropriately modified clones. Various specific amino acid sequences can be encoded that have unique affinities for agents that will permit efficient protein purification or will provide binding properties so that proteins of interest can be immobilized in large arrays such as on micro chips. These arrays will be available to the user community and could be used by them or within the facility to screen for molecules that will specifically detect each protein. These reagents will then be invaluable to the larger community for detecting and measuring individual protein products in order to locate them within molecular machines, within cells, and within communities of microorganisms in both experimental and environmental samples.

Critical reagents for protein identification and versatile biological experimentation will also be produced routinely. These include such reagents as polyclonal and monoclonal antibodies for the protein product samples to permit tracking the protein locations *in vitro*, in whole cells, and tissue single molecule and ensemble imaging microscopy.

FACILITY INFRASTRUCTURE REQUIREMENTS

The GTL-1 facility has a number of specialized design requirements that are unlikely to be met in a cost effective manner by any existing physical space that could be leased. The facility needs to support a diverse set of activities, including:

- Robotics for high-throughput production and characterization
- Mass spectrometry instrumentation with specialized physical requirements
- Other specialized instrumentation for protein characterization
- Clean rooms for PCR
- Large scale fermentation for producing proteins in quantities as large as hundreds of milligrams
- Large freezer capacity for storage of clones, proteins, tags
- Large computer facility to handle the bioinformatics activities
- Large power grid to support all the computer and instrumentation needs
- HVAC that can handle the heat load from all the above activities, as well as ventilation for chemistry operations
- High roll up doors for bringing in and out large equipment

The Subcommittee extensively discussed the question of a centralized vs. decentralized GTL-1 facility. Within these bounds there were several thoughts about options for the relationships between the individual components, and recognition that some degree of distributed research activity was anticipated. The motivation for centralization was driven by practical issues; including the utility of a specifically designed building and the economy of scale that invariably occurs when a complex set of activities occurs under a single roof. The facility will need to be relatively large, housing multiple robots, liquid handling devices, reagent storage, fermenters, sequencers, NMRs, mass spectrometers, other spectroscopic instrumentation and other specialized tools for protein classification. It will also require a substantial computer hardware facility to support all the informatics, storage and LIMS requirements. The operation and maintenance of all these in an optimal manner is best achieved by bringing together the components in one place, where the resident expertise is high and is optimized to the very challenging demands of a very high-throughput, economical operation.

Preference of the Subcommittee for centralization was also motivated by the recognition that the collective housing of unified and dedicated management and scientific staff would enable strong synergism. The vision of the facility is that it will integrate a stream of production that satisfies the basic mission of delivering expressed proteins to end users in the GTL science program, while simultaneously developing and integrating novel and incremental technologies that will further the specific mission of the group. This is a demanding portfolio and requires a first rate scientific and intellectual environment.

This centralized model has been well proven in the current DNA sequencing centers with which DOE BER has significant experience at its very successful Production Sequencing Facility in California. In such a sequencing center, expertise is developed around process integration, automation and high levels of equipment capitalization. The advantage of a single centralized facility is felt across all areas, including everyday management and process development, and has validated that this model is most effective. To the extent that the GTL-1 facility has an analogous mission to the DNA sequencing centers, the benefits of centralization are clear. The component of the GTL-1 that is identical to the sequencing centers is the activity that focuses on the improvement and scaling of routine production of individual proteins, with an accompanying decrease in cost and increase in quality.

The most important difference between the GTL-1 and DNA sequencing centers is that the protein production technologies, and the proteins themselves, do not uniformly lend themselves to easy, scaled production. As a consequence, the GTL-1 facility will need to be deeply immersed in practical R&D issues, including, for example, how to express and harvest membrane bound proteins, how to co-express factors that will enable better solubility, and how to better purify large protein complexes.

Hence, the Subcommittee concludes that while existing space may meet some of the requirements, the comprehensive set of activities that must be supported for GTL-1 to achieve its goals requires a specific design process that incorporates input from the operators and users. It would be extremely costly and ineffective to develop this with existing leased space. Given the anticipated lifetime of the GTL-1 facility (15-20 years or more), leased spaced is not likely to be cost effective.

In discussing the need for GTL-1 R&D, the Subcommittee discussed where this kind of effort should occur. Two broad areas were discussed: 'research' that was highly innovative and potentially enabled great increases in efficiency for the facility; and 'development', that was geared towards more incremental improvements in the protein production or incorporating 'research' developments into the production pipeline. These areas overlap considerably, but the Subcommittee felt that while a centralized facility must have the resources to carry out the 'development' components, the most innovative approach to the 'research' component was to distribute it in such a way to engage a broad range of effort from outside of the actual facility.

One specific example of the need for a strong 'research' component is well illustrated by the nature of protein modification after proteins are synthesized. Heterologously expressed proteins are not guaranteed to be biologically relevant given that various types of post-translational modifications (PTM) may not be affected by a given expression system. Since a major goal of the GTL-1 facility is functional characterization of proteins, obtaining information about PTM from the original organism and producing fully functional proteins is critical. There is no current technology that the Subcommittee is aware of to do this in a high-throughput manner.

These distributed 'research' activities could occur at other national laboratory locations, in universities or in industrial laboratories – and could be funded under the core program or by other mechanisms. In any case, the fruits of the distributed activities will need to be accessed by the GTL-1 facility management. Apart from the usual mechanisms of scientific communication, the GTL-1 facility will have an incentive to seek the results of the 'research' efforts, driven by its internal 'development' program and goals of continued improvement in efficiency and quality in production.

Lastly, given the rapid development of biotechnology, it is imperative that the design recognizes the need for the facility to be flexible and evolvable. This is in itself a key design requirement that will allow the introduction and substitution of new technologies that will be added to GTL-1, for example, by incorporating modular work spaces and large crawl spaces to accommodate economical refits as needed.

LIGHT SOURCE CO-LOCATION

The Subcommittee considered the question of "Does the GTL-1 facility have to be next to a synchrotron light source (SR) or not?" The following account reports the process and rationale that led to the conclusion that GTL-1 facility need not be located in close proximity to a SR.

The framework for this deliberation was a discussion of methodologies needed for protein characterization that would be facilitated or enhanced by SR. Foremost among these is macromolecular single crystal x-ray diffraction, where it has been very clearly proven that the high brightness, collimation and tunability of SR are key properties that enable study of the most complex and challenging problems in structural biology. Further, SR is essential for high-throughput operation as is currently being developed for structural genomics. Small angle x-ray scattering is also a technique that is much better performed on biological systems with SR. SAXS provides lower resolution structural information, which can be valuable in the study of large protein complexes and can also be used to investigate protein folding state and folding dynamics. A third area of

potential importance is circular dichroism (CD) spectroscopy where SR offers advantages in flux in a shorter wavelength regime than that available from conventional sources (120-180 nm). Contained in this region are spectral features that are particularly sensitive to secondary structural elements in proteins.

With this background information, the Subcommittee then considered the question of the need for SR proximity and the relative importance of the methodologies in the context of the GTL-1 facility. Macromolecular crystallography (along with NMR for proteins whose size allows them to be studied) was viewed as the primary tool that would be needed to obtain atomic or near atomic resolution structural information. While such information may not strictly be needed for "characterization", it certainly is an essential component of understanding function. However, it has now been widely demonstrated that transporting frozen crystals and utilization of robotic-enabled beam lines at the SR facilities is both an effective method of operation and greatly enables efficiency of operation and high-throughput. Immediate proximity to the actual SR source is certainly no longer an important factor in this class of experiments. While there are some exceptions (*e.g.*, very large complexes that typically give fragile, weakly diffracting crystals), it was felt that on balance, remote access was capable of meeting the majority of needs.

SAXS is performed on proteins in solution. There is certainly valuable information that could be obtained on protein folding state and folding pathways/kinetics but it is the opinion of the Subcommittee that, given present knowledge, this is unlikely to be required in a high-throughput, large volume screening mode for GTL-1. Thus while synchrotron-enabled SAXS could be a valuable methodology among those available for characterization of the proteins being produced by GTL-1, this could again be managed by transporting the samples to the facility. Immediate information on folding state could be obtained by other techniques like light scattering. The value of CD spectroscopy in high-throughput structural characterization was somewhat less clear to the Subcommittee as the shorter wavelength applications are not as mature as with the x-ray based methods. However, it was felt that given current knowledge, SR-enabled CD also did not offer a compelling reason to be in close proximity to a synchrotron source.

COMPLEMENTARITY TO PROGRAMS IN OTHER AGENCIES

The Subcommittee discussed the relationship of the DOE BER GTL science program and the GTL facilities (especially GTL-1) to the efforts of other agencies. Members of the Subcommittee were most knowledgeable about the Protein Structure Initiative (PSI), a program in the National Institutes of General Medical Sciences (NIGMS) at NIH and thus focused on this comparison. The PSI initiative has several components, but that most comparable to aspects of the GTL science program and the GTL-1 facility is the program in structural genomics (SG). The NIGMS SG program, currently nearing the end of a 5-year pilot phase, has been focused on the development of high-throughput pipelines to go from gene sequences to three dimensional protein structures. High-throughput, automated approaches are one focus of the R&D and are viewed as required to bring the cost per structure down to a manageable level. The pilot program, involving 9 grants to create SG centers, has been extremely successful and NIGMS has requested proposals to establish SG centers in phase II of the program (anticipated to be about a 5-year effort with a total investment of about $375M). NIH anticipates funding 6-11 centers in which there are two types - production and

technology development (with the later focusing on challenging problems such as protein complexes and membrane structures). Evaluation of the grant applications is currently underway and awards are expected to be announced later in FY2005.

The PSI SG program is an excellent example of a distributed approach that focuses on a very specific target (a large number of new 3D, high resolution structures to gain a detailed and predictive knowledge of protein structure). Protein production is one of the central aspects of this effort since it produces the material required for crystallization trials and structure determination. Thus there are parallels between the PSI initiative and the planned GTL-1 facility and there are certainly areas where development and sharing of new technologies will benefit both initiatives significantly. While estimates on production goals of the PSI-II initiative vary depending on assumptions like the number of expressed/purified proteins needed to yield a new structure, a reasonable expectation for the protein sample production is the order of 10,000 – 15,000 per year over all the centers. The most frequently stated goal of GTL-1 is about 10,000 proteins produced and characterized as to their purity and properties (not all with 3D structures) per year. Hence to first approximation, the production goals of the two initiatives are in the same order of magnitude. The Subcommittee believes that both approaches offer viable strategies but there are programmatic considerations that influence the recommendation of a centralized GTL-1 facility.

The NIGMS PSI initiative lends itself very well to being a distributed effort. Coordination is required for target selection but the goal of a large number of new structures can be reached without consideration of the specific type of organism selected for the protein source (ranging from human, mouse or other mammalian to various organisms). In contrast, the BER GTL-1 facility will target specific microbes that are mission relevant to DOE goals and produce those proteins and reagents that are important for the GTL science program investigators. For each targeted microbe, however, GTL-1 has the more ambitious goal of producing the entire proteome in order to use this combined information for DOE missions. However, the PSI centers that focus on specific protein types will be important sources of new technologies for high-throughput production of many proteins, and thus will be an excellent resource for GTL-1. The focus on microbial biology is indeed one of the main defining features of the GTL science program as enabled by the GTL-1 facility. On balance, the Subcommittee feels that this can best be accomplished by a centrally operated and managed facility.

SUBCOMMITTEE MEMBERS

Dr. Raymond F. Gesteland (Chair, Subcommittee)
University of Utah

Dr. Keith O. Hodgson
Stanford University

Dr. Marvin Cassman
San Francisco, CA

Dr. David Eisenberg
UCLA

Dr. Jim K. Fredrickson
Pacific Northwest National Laboratory

Dr. David J. Galas
Keck Graduate Institute

Dr. Richard A. Gibbs
Baylor College of Medicine

Dr. Eric Mathur
Diversa Corporation

Dr. Abigail Salyers
University of Illinois, Urbana-Champaign

Dr. Robert A. Scott
University of Georgia

Dr. Lisa Stubbs
Lawrence Livermore National Laboratory

Dr. Matthew Tobin
Codexis, Inc.

Dr. Jill Trewhella
University of Utah

Dr. Barbara J. Wold
California Institute of Technology

Dr. John C. Wooley
University of California, San Diego

Participating by phone:

Dr. Mike Snyder
Yale University

Dr. Claire Fraser
The Institute for Genomic Research

Dr. James Tiedje
Michigan State University

Dr. Janet Smith
Purdue University

Dr. Jonathan Greer
Abbott Laboratories

Dr. David Burgess
Boston College

CHARGE LETTER:

Department of Energy
Office of Science
Washington, DC 20585

Office of the Director

September 23, 2004

Dr. Keith O. Hodgson
Director, Stanford Synchrotron Radiation Laboratory
Department of Chemistry
Stanford University
Stanford, CA 94305

Dear Dr. Hodgson: Keith

By this letter, I am charging BERAC to provide advice to the Office of Science on the potential scientific impact of the proposed Facility for the Production of Proteins and Molecular Tags (GTL-1), and some of the major design characteristics essential to its effectiveness.

In its "Report on the Biological and Environmental Research Advisory Committee's (BERAC) Consideration of the Genomes to Life Facilities Plan", BERAC stated that, "These new facilities need to be brought into operation and be effective if GTL itself is to succeed as a science program." I request that BERAC address the specific impact of GTL-1 on the GTL program.

Questions to be addressed:

1. Why is this facility essential to meeting the scientific goals of the program?
2. Is the facility described the right one in terms of output/capabilities/dimensions to meet the goals?
3. Does it have to be a central facility or can it be distributed?
4. As described does it require a specially designed building or can the needs be met with leased space, equipment, etc.?
5. What scientific probes, such as synchrotron light sources or mass spectrometers, are required to make the facility successful?
6. Does it have to be next to a light source or not?

Finally, recognizing the interrelationship between biology research programs funded at all Federal agencies, I ask that you comment on the role of, interests in, or impacts of GTL-1 on the biology research programs of other agencies.

Printed with soy ink on recycled paper

I request that BERAC report on its findings and recommendations at its November 3-4, 2004 meeting.

Sincerely,

Raymond L. Orbach
Director